The Prevalence and Cognitive Profile of Sequence-Space Synaesthesia

Jamie Ward, Alberta Ipser, Eva Phanvanova, Paris Brown, Iris Bunte and Julia Simner

School of Psychology, University of Sussex, Brighton, UK

Manuscript Submitted to: Consciousness and Cognition

Running Head: sequence-space synaesthesia

Address for Contact:

Prof. Jamie Ward,

School of Psychology,

University of Sussex,

Falmer, Brighton,

BN1 9QH, U.K.

Tel. : +44 (0)1273 876598

Fax. : +44 (0)1273 678058

E-mail : jamiew@sussex.ac.uk

People with sequence-space synaesthesia visualize sequential concepts such as numbers and time as an ordered pattern extending through space. Unlike other types of synaesthesia, there is no generally agreed objective method for diagnosing this variant or separating it from potentially related aspects of cognition. We use a recently-developed spatial consistency test together with a novel questionnaire on naïve samples and estimate the prevalence of sequence-space synaesthesia to be around 8.1% (Study 1) to 12.8% (Study 2). We validate our test by showing that participants classified as having sequence-space synaesthesia perform differently on lab-based tasks. They show a spatial Stroop-like interference response, they show enhanced detection of low visibility Gabor stimuli, they report more use of visual imagery, and improved memory for certain types of public events. We suggest that sequence-space synaesthesia develops from a particular neurocognitive profile linked both to greater visual imagery and enhanced visual perception.

Keywords: sequence-space, synaesthesia/synesthesia, diagnosis, perception, imagery, memory.

Introduction

People with sequence-space synaesthesia (SSS) visualize sequential concepts such as numbers and time (e.g. years, months) as an ordered pattern (or 'spatial form') extending through space. These may be complex (e.g. undulating, spiraling) or simple patterns (linear); three-dimensional or two-dimensional; projected externally (e.g. as a hoop around the body) or viewed on some internal 'inner screen' (Eagleman, 2009; Sagiv, Simner, Collins, Butterworth, & Ward, 2006; Smilek, Callejas, Merikle, & Dixon, 2007). They can sometimes be highly prolific extending to sequenced concepts such as temperature, weight, shoe sizes, etc. (Hubbard, Ranzini, Piazza, & Dehaene, 2009). The prevalence and nature of this form of synaesthesia, and its links to cognitive ability, remains an enduring debate extending back to the Nineteenth century (e.g. Calkins, 1895; Galton, 1880a). One reason why the debate is still unresolved is that, compared to other types of synaesthesia (e.g. where the unusual experience is colour), there remains no commonly agreed diagnostic measure for SSS. In this set of studies, we take a significant step towards resolving this debate by further developing and validating a recently devised diagnostic test (Rothen, Jünemann, Mealor, Burckhardt, & Ward, 2016), and using it to assess the prevalence and cognitive profile of the (probable) synaesthetic group that pass it.

The Victorian polymath, and cousin of Charles Darwin, Francis Galton initially became interested in SSS because of its possible link to the familial inheritance of mental ability (Burbridge, 1994). Galton's interest began when he read an obituary of a famous calculating prodigy and engineer, George Bidder (1806-1878), written by his son in which it was noted that both father and son had unusual visual imagery abilites. This included a 'number form' (as Galton called it) drawn by the son together with several forms depicting time (months, historical years). Through extensive surveys of other people, Galton concluded that these forms are a particular kind of mental imagery, created during childhood, that becomes more vivid and automatic in adulthood in those that make use of it but that disappears in those who do not (Galton, 1880b, 1880c). He estimated (without firm empirical evidence) that they were more prevalent in women and children, and he gave these estimates as 25% in school boys, 6.7% in adult females and 3.3% in males (Galton, 1880a, 1880c). Galton was unaware of the emerging literature on synaesthesia in continental Europe (Jewanski, Day, Simner, & Ward, 2011) and did not draw a link to these cases, although others soon did (e.g. Flournoy, 1893).

Other early estimates of the prevalence of SSS in the general population are 16.7% (Patrick, 1893), 11.1% (Flournoy, 1893), 12% (Calkins, 1895), and 7.3% (Phillips, 1896-97). In the contemporary literature, estimates include 14.2% (Seron, Pesenti, Noel, Deloche, & Cornet, 1992) and 12% (Sagiv et al., 2006) and 4.4% (Brang, Teuscher, Ramachandran, & Coulson, 2010). Most of these estimates tend to be based around number forms (i.e., number-space synaesthesia) although, when noted, similar prevalences tended to be found for calendars and the same individuals tend to report more than one spatial form. Typically, to meet the criteria for these studies one has to both verbally confirm the presence of a spatial form (e.g. “Do you think about numbers [letters/days/months] as being arranged in a specific pattern in space?” from Sagiv et al., 2006) and produce a drawing of it - which in some studies was also judged to be consistent over time. However, these criteria are not always as stringent as those used to diagnose other forms of synaesthesia and it is unclear how well these methods discriminate between those with true SSS and those who visualise similar representations only during testing as a result of being prompted by the experimenter, and/or rely on simple reproductions of patterns in the environment (e.g. layouts in calendars). Several studies have used Stroop-like interference tests to show that people with SSS behave differently insofar as their spatial associations are more automatic. For instance, making left/right responses to indicate the placement of a number or month on-screen is slower if the position of the stimulus is incongruent with respect to the synaesthesia (e.g. Sagiv et al., 2006; Smilek et al., 2007). These studies provide support for the authenticity of SSS (at the group level), but the interference effect at the individual level can often be small and variable making it less useful as a diagnostic measure.

For synaesthesia involving colour, the current standard is to use a computerized colour picker to select a colour for a stimulus and repeat the procedure several times so that a consistency score can be calculated (the average distance between colour selections for the same items; Eagleman, Kagan, Nelson, Sagaram, & Sarma, 2007; Rothen, Seth, Witzel, & Ward, 2013). Several tests have been developed for sequence-space synaesthesia based on the same principle, with consistency measured as the distance in some 2D or 3D space (e.g. Brang et al., 2010; Eagleman, 2009). However, these have been limited by the absence of normative cut-offs for diagnosis, and have sometimes been limited to those with particular characteristics (e.g. those who visualise their form in peripersonal space, Smilek et al., 2007). Rothen et al. (2016) attempted to address these shortcomings. They asked people with SSS to reproduce their spatial form on a 2D computer screen by making mouse clicks to indicate where each item in the sequence should be placed spatially and repeating this three times (people with 3D representations are generally able to represent them in 2D). Those without SSS were asked to think about numbers and time (days, months) spatially but were given no particular instructions as to how to do this. Using a variety of different measures of consistency Rothen et al. determined the optimal way of discriminating between the groups (based on the area bounded by their selections) and suggested a cut-off for diagnostic purposes. The current study extends this measure by creating the test online (rather than in-person) and developing a questionnaire to replace self-report. By running the measure on large samples of naive participants (i.e. not recruited on the basis of having synaesthesia) we aim to determine the prevalence and also determine whether SSS identified by these means have particular cognitive abilities in imagery, perception, or memory.

Contemporary research on SSS has provided more direct support for Galton's (1880a) proposal that it is related to the phenomenon of mental imagery. People with SSS tend to self-report more vivid mental imagery (e.g., Price, 2009; Spiller, Jonas, Simner, & Jansari, 2015). They also perform better on cognitive tests of mental imagery ability (e.g. Brang, Miller, McQuire, Ramachandran, & Coulson, 2010; Havlik, Carmichael, & Simner, 2015; Simner et al., 2009; but see Rizza & Price, 2012). This superiority is linked particularly to those SS synaesthetes who mentally project sequences into external space versus the mind's eye (Havlik et al. 2015). There is also evidence that people with SSS have enhanced visual perception abilities. Ward, Rothen, Chang and Kanai (2017) administered a battery of visual perception tests to synaesthetes with grapheme-colour synaesthesia, sequence-space synaesthesia, or both. The synaesthetes as a whole were better at both colour discrimination and a measure of shape discrimination (but they did not have better visual ability in a global sense). The SSS groups, relative to controls and grapheme-colour, also had an advantage at detecting low visibility grating stimuli (Gabor patches) particularly at high spatial frequencies. The explanation for this is not fully clear, but the suggestion is that differences in perceptual sensitivity may be a prerequisite for the development of synaesthesia (Shriki, Sadeh, & Ward, 2016). Finally, SSS might be linked to memory ability. For example, spatial representations of time (years, months, days) might be used to structure memories for events – a mnemonic strategy that is not available to others. Simner, Mayo and Spiller (2009) found that people with SSS were better able to date news and cultural events than others, and could generate more autobiographical events given a year cue than others. In summary, whilst there is good evidence that SSS is linked to certain cognitive abilities in imagery, perception and memory, a question-mark still hangs over all these findings given that there is no agreed basis for diagnosing SSS. For instance, previous results would be confounded if it turned out to be very hard to discriminate people with SSS (who have good mental imagery) from people who have good mental imagery but don't have SSS.

In the two studies below we use the diagnostic consistency test of Rothen et al., (2016) to determine the prevalence of SSS in individuals who were not recruited on the basis of having synaesthesia and who were not aware that the testing was about the condition. We also develop a novel questionnaire, to replace self-referral as the subjective measure. Thus, to be classed as having SSS participants must both report its subjective presence (via questionnaire) as well as meet an objective criterion (via consistency test). In Study 1 we compared the naïve sample against a set of people with probable SSS (who are active amongst the synaesthesia community). In Study 2, we recruit a new sample of naïve participants using the same diagnostic test (consistency and questionnaire) and assess them on a battery of tests that have previously been claimed to be relevant to SSS. The validity of the diagnostic test lies in its ability to predict other cognitive traits not directly linked to the test itself (otherwise it is entirely circular). The tests included: questionnaire measures of mental imagery, personality, and cognitive style; a spatial Stroop measure of automaticity; the ability to detect low visibility Gabor stimuli; memory for public and autobiographical events; and a blind-scored interview about visualizing numbers and time. The ability to accurately diagnose SSS is not simply a matter of developing and validating new tests but, rather, it goes to the heart of the theoretical debate as to what the nature of synaesthesia is. Is synaesthesia an unusual variant of mental imagery (Galton, 1880a)? Is it a memory phenomenon, e.g. relating to the learning and retention of associations (e.g. Yon & Press, 2014)? Does it depend strongly on mechanisms of visual perception (e.g. Ramachandran & Hubbard, 2003)? We return to these important questions in the Discussion, in light of our findings.

Study 1: Prevalence and Characteristics of Sequence-Space Synaesthesia in a Large Online Sample

Method

Participants

There were two groups of participants. The first group was a set of self-declared sequence-space synaesthetes who are active members of an online synaesthesia forum (http://www.daysyn.com/Synesthesia-List.html) and were recruited via this site. The self-declared SSS group comprised 27 participants (mean age = 35 years, SD = 17.8; 18 female, 6 male, 3 undisclosed). The second group of participants took part in the prevalence study. They were recruited online via Amazon's online crowdsourcing marketplace Mechanical Turk, without any explicit mention of synaesthesia in the study information (the study was called "Numbers, Time and Space"). MTurk has been shown to generate reliable data and a more representative demographic sample than commonly used in psychological research (Buhrmester, Kwang, & Gosling, 2011). The MTurk sample consisted of 419 participants (195 females, 223 males and 1 undisclosed gender), aged between 18 and 73 years (M = 37.00, SD = 12.70). Participants were paid $0.70.

This study was approved by the Cross-Schools Science and Technology Research Ethics Committee at the University of Sussex and the study was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Procedure

All participants were tested remotely via an online test created in Inquisit (www.millisecond.com). The study contained two parts: participants first completed a test of spatial consistency used by Rothen et al. (2016), which was then followed by a novel questionnaire aimed at quantifying the subjective presence of SSS. On average the study took around 15 minutes to complete.

For the consistency test, participants were given the following instructions:

> "In this study you will see single digits (e.g. 5), days of the week (e.g. Tuesday), and months of the year (e.g. July) displayed in the centre of the computer screen. Your task is to think about how these concepts may be arranged spatially on the 2D computer screen. Some people may automatically think about these concepts spatially in their everyday life, and you should use this. For other people this may seem like a strange task, but just go with your intuitions. There isn't a right or wrong answer. When you see each item on the computer screen, then visualise where it fits spatially and click the mouse in the corresponding location on the screen (using a left click). Each item is repeated three times."

A number (digits 0-9), day (N=7) or month (N = 12) was presented for 1 s in the centre of a white screen followed by a fixation cross that remained visible until the participant responded with a mouse click. The next stimulus appeared immediately after the click. Each number, day and month was presented three times. The order of the stimuli was randomised, but such that no stimulus was repeated until the previous batch of unique stimuli (N = 29) had been presented. The font was Courier New, size 18, and in bold typeface. Although it is not possible to control for viewing conditions, Inquisit enables the monitor resolution to be logged. The mode was 1366x768, and the lowest and highest monitor resolutions were 800x600 and 2560x1440. We establish later that the results of this online study are comparable to those used on a standard monitor in laboratory settings reported by Rothen et al. (2016).

The questionnaire consisted of 15 questions (Appendix 1). Two questions related to age and gender. Nine questions were answered on a 5-point Likert scale (ranging from Strongly Agree to Strongly Disagree). An example is: "Some people routinely think about sequences as arranged in a particular spatial configuration, do you think this might apply to you?" This question was accompanied by two diagrams of spatial forms by sequence-space synaesthetes. Four of the nine Likert questions were reverse coded (i.e. we would expect someone with SSS to disagree with such statements). Three questions asked about qualitative aspects of the spatial form (location in space; characteristics such as font and perspective; whether found for temperature, weights, etc.) and a final question invited open-ended comments about the task and strategies used.

Analysis

For the consistency test, the data consist of three xy coordinates for each stimulus. For example, there would be three xy coordinates for the month January, and these correspond to the spatial location chosen for January each time it was presented on-screen. Rothen et al. (2016) compared several objective measures of consistency for discriminating SSS from controls and, using a quantifiable method based on receiver operating characteristics (ROC), found that the area of the triangle given by the three coordinates offers the best sensitivity and specificity at a calculated cut-off of 1596 pixels. That study used a standard monitor of 1024x768 and this equates to an average triangle area of <0.203% of the monitor size as diagnostic of SSS. Given the variability in monitor size in the present study, the <0.203% cut-off was used rather than number of pixels. One problem with this measure is that participants can obtain a low score by trivially clicking on the same spatial location (e.g., the centre of the screen) for every trial. Whilst this did not occur for any of our self-declared sequence space synaesthetes it did occur in about a quarter of the MTurk sample. Any scores <.203 were checked to see if such a strategy was used and those participants were reclassified as failing the consistency test (but not discarded from the study).

For the questionnaire measure, the nine questions using the Likert scale were summed together. Low scores indicate a high level of SSS. Given that each question may be given a score of 1-5, the minimum score was 9 (an ideal synaesthete) and the maximum score was 45 (an ideal non-synaesthete). The results consider several possible cut-offs for this measure. The reliability of this scale, based on the MTurk sample, was good (Cronbach's alpha = .885) and was not improved by deleting any items.

Results

The self-declared sequence-space synaesthetes, recruited via the online synaesthesia forum, scored an average of 0.15% (SD = 0.13) on the spatial consistency test. This is comparable to that reported by Rothen et al. (2016) in standardised laboratory conditions (M = 0.14%, SD = 0.17). As such, we are able to confirm that the translation of this test to an online platform with non-standardised viewing conditions has not compromised the integrity of the task. As the questionnaire was entirely new, there was no pre-existing cut-off for this measure. We remind the reader that low scores indicate high levels of report of SSS. The self-declared synaesthetes (recruited by our a priori targeted invitation to the synaesthesia community) had a mean score of 13.67 (SD = 5.09; range = 9-29). This provides some constraints on what would be a reasonable range of cut-offs for the questionnaire: with a score of 17 being the 75th centile and 21.5 being the 95th centile.

In order to determine the prevalence of SSS, we ascertained the proportion of people in the naïve MTurk sample who both behave like a sequence-space synaesthete (on the consistency measure) and claim to have sequence-space synaesthesia (on the questionnaire measure). Figure 1 shows the prevalence of SSS, in the grey cells, within our naïve sample based on a cut-off of <.203 in the consistency test (achieved using a non-trivial strategy), and based on four different cut-offs from 18 to 21 in our questionnaire (low scores indicate high levels of SSS). The prevalence ranges from 7.2% (≤18; i.e. a more conservative threshold) to 9.5% (≤21; i.e. a less conservative threshold). It is to be noted that increasing the consistency cut-off by up to 50% (to <.300%) has only modest effects on the prevalence estimate (9.5% at ≤18 and 12.6% at ≤21). We suggest that any cut-offs in this recommended range (18-21 for questionnaire and 0.2-0.3% for consistency) will produce two groups that are predominantly made up of SSS and non-SSS that are sufficient to generate reliable group differences on other measures that are sensitive to SSS – as we show in Study 2. For the remainder of this paper, we will adopt the more conservative cut-off of ≤19 on the questionnaire and <.203 on the consistency test.

		Consistency	
		Pass	Fail
Qu'aire ≤18	Pass	7.2	9.5
	Fail	13.9	69.4

		Consistency	
		Pass	Fail
Qu'aire ≤ 19	Pass	8.1	11.5
	Fail	14.9	65.5

		Consistency	
		Pass	Fail
Qu'aire ≤ 20	Pass	9.1	14.6
	Fail	15.9	60.4

		Consistency	
		Pass	Fail
Qu'aire ≤ 21	Pass	9.5	17.2
	Fail	16.7	56.6

Figure 1: The estimated percentage prevalence of sequence-space synaesthesia (grey cells) as a function of different questionnaire ("Qu'aire") cut-offs (from top left: ≤18 to ≤21). The other cells display participants who fail one or both measures.

Irrespective of the precise cut-offs, the largest group of participants are those who claim not to have sequence-space synaesthesia and do not pass the consistency test. However, there are two groups for whom the objective test (consistency) and subjective test (questionnaire) do not concur. Some participants pass the consistency test but deny having synaesthesia. Our speculation about this group is that they have developed an effective strategy of visualising numbers and time <u>during the task</u>, but do not have habitual (and automatic) spatial forms. Another group of participants fail the consistency test but do claim to have SSS on the questionnaire. This group could include some genuine synaesthetes who, for whatever reason, have difficulty in generating a highly consistent form on our test (e.g. some attempt to drag the word from the centre rather than click on the location). But we speculate that the larger portion of this group did not answer the questions with careful consideration.

What are the characteristics of the people who we have identified as having SSS and what are the characteristics of their spatial forms? These are summarised in Table 1 along with the characteristics of our self-referred sample. In the MTurk sample, the prevalence for males (8.5%) and females (7.7%) did not differ significantly ($\chi^2(1)=.758$, p = .384). This stands in contrast to the high female:male ratio in our self-referred sample (3:1). However, a female bias in non-naïve samples, compared to no female bias in naïve opportunity sampling is exactly what has been observed in grapheme-colour synaesthesia (e.g., Simner & Carmichael, 2015). The targeted synaesthete group and the naïve prevalence groups also differed in a number of other respects: the former were more likely to report colours and shading, and for them to be more 3D. In most other respects the groups tended not to differ. For instance, both report a prolific number of forms (not just for the ones tested here) and tend to experience them internally rather than projected externally.

Table 1: Qualitative characteristics of the spatial forms in the targeted synaesthete group, and the candidate synaesthetes identified from the naïve prevalence group with MTurk (questionnaire≤19, consistency<.203%). NA: Where expected values < 5, χ^2 statistics are not reported.

	Self-Referred SSS Group	**Prevalence Study SSS**	**Difference between Groups**
N	27	34	
Female:Male ratio	3:1	0.9:1	**$\chi^2(1) = 5.471$, $p = .019$**
Which forms reported? (%)			
Numbers	81	82	$\chi^2(1) = 0.008$, $p = .930$
Days	93	94	NA
Months	100	97	NA
Years	89	71	$\chi^2(1) = 3.006$, $p = .083$
Alphabet	81	79	$\chi^2(1) = 0.041$, $p = .840$
Temperature	59	62	$\chi^2(1) = 0.108$, $p = .742$
Height	48	50	$\chi^2(1) = 0.021$, $p = .886$
Weight	44	47	$\chi^2(1) = 0.041$, $p = .839$

Spatial location (%)			
Outside body	37	9	NA
Inner screen	56	68	NA
Inside body	7	21	NA
Characteristics (%)			
Colours	37	3	$\boldsymbol{\chi^2(1) = 12.201, p < .001}$
Shading	26	3	NA
2D	41	65	$\chi^2(1) = 3.481, p = .062$
3D	52	9	$\boldsymbol{\chi^2(1) = 8.854, p = .003}$
Perspective	33	32	$\chi^2(1) = 0.007, p = .935$
Like blocks or tiles	33	26	$\chi^2(1) = 0.341, p = .559$
A certain font	7	3	NA

In summary, we estimate the prevalence of sequence-space synaesthesia to be around 8.1% based on a combination of high spatial consistency, and questionnaire self-report (scores ≤19). This may be a lower limit given that we assess people who report SSS for both time and number. These naïve individuals found in our test resemble in many ways (but not all ways) self-declared synaesthetes actively participating in the online synaesthesia community.

Study 2: Prevalence and Characteristics of Sequence-Space Synaesthesia in a University Sample

Study 2 adopts a similar screening approach on a University sample, but follows up with a detailed battery of in-person tests on a subset of participants. The tests were chosen because they have previously been reported to be related to sequence-space synaesthesia. However, this previous research has not always used a formal method of diagnosis. Specifically, we hypothesise that sequence-space synaesthesia will be linked to a distinct cognitive style reported by Mealor et al. (2016): namely increased imagery ability, increased technical/spatial imagery, increased systemising, increased interest in language, and reduced global bias (i.e. more local bias), all measured using the Sussex Cognitive Styles Questionnaire. We hypothesise that sequence-space synaesthesia will show Stroop-like interference when months are presented in incongruent spatial locations. We hypothesise that they will show enhanced ability to detect low visibility Gabor stimuli (following Ward et al., 2017) and enhanced memory for dating of events and in recalling autobiographical memories (following Simner et al., 2009). Finally we test whether any differences across groups might be attributed to personality traits by administering the Ten Item Personality Questionnaire

(TIPI) (Gosling, Rentfrow, & Swann, 2003). Previous studies suggest that synaesthesia might be linked to a distinct personality profile. For example, Rouw and Scholte (2016) showed that self-reporting synaesthesia was linked to higher scores on Openness to Experience and Neuroticism and lower scores on Conscientiousness, relative to controls. These findings overlap with those of Banissy, Holle, et al. (2013) whose sample of verified grapheme-colour synaesthetes scored higher on Openness to experience and lower on Agreeableness. Chun and Hupe (2016) also reported that their verified synaesthetes scored higher on Openness, as well as a related trait of absorption, which is an individual's' participation in and enjoyment of imaginative activities. If all variants of synaesthesia share the same personality profile, we might predict that synaethetes in the current study will score higher on the same traits previosuly reported to be linked to the condition, such as as openness to experience.

METHODS

Participants

Our initial sample consisted of 235 participants (37 males; M = 20.13, SD = 3.53, Range = 18-49 years). The sample was recruited from the undergraduate student population of the University of Sussex, who took part for course credit. These participants completed an online study primarily to enable us to determine who is likely to have sequence-space synaesthesia.

All participants in the SSS category were then invited to participate in the second session of in-person testing, and N = 13 completed the second session (2 males; Mean age = 22.6 years, Range = 18-33). In this second session we additionally recruited N = 51 Non-SSS controls (12 males; mean age = 20.6 years, range = 18-44). These controls had failed either the SSS consistency test (N = 9), the SSS questionnaire (N = 14), or both (N = 28). For the second session, all participants were paid either £10 or received course credits. The study was approved by the University of Sussex Cross-Schools Sciences and Technology Research Ethics Committee and the study was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. All participants provided their consent to participate in the study.

Materials and Procedure

The overall procedure and timeline of testing is summarised in Figure 2. The tests were always administered in this order, and the online session and in-person session were always at least one week apart.

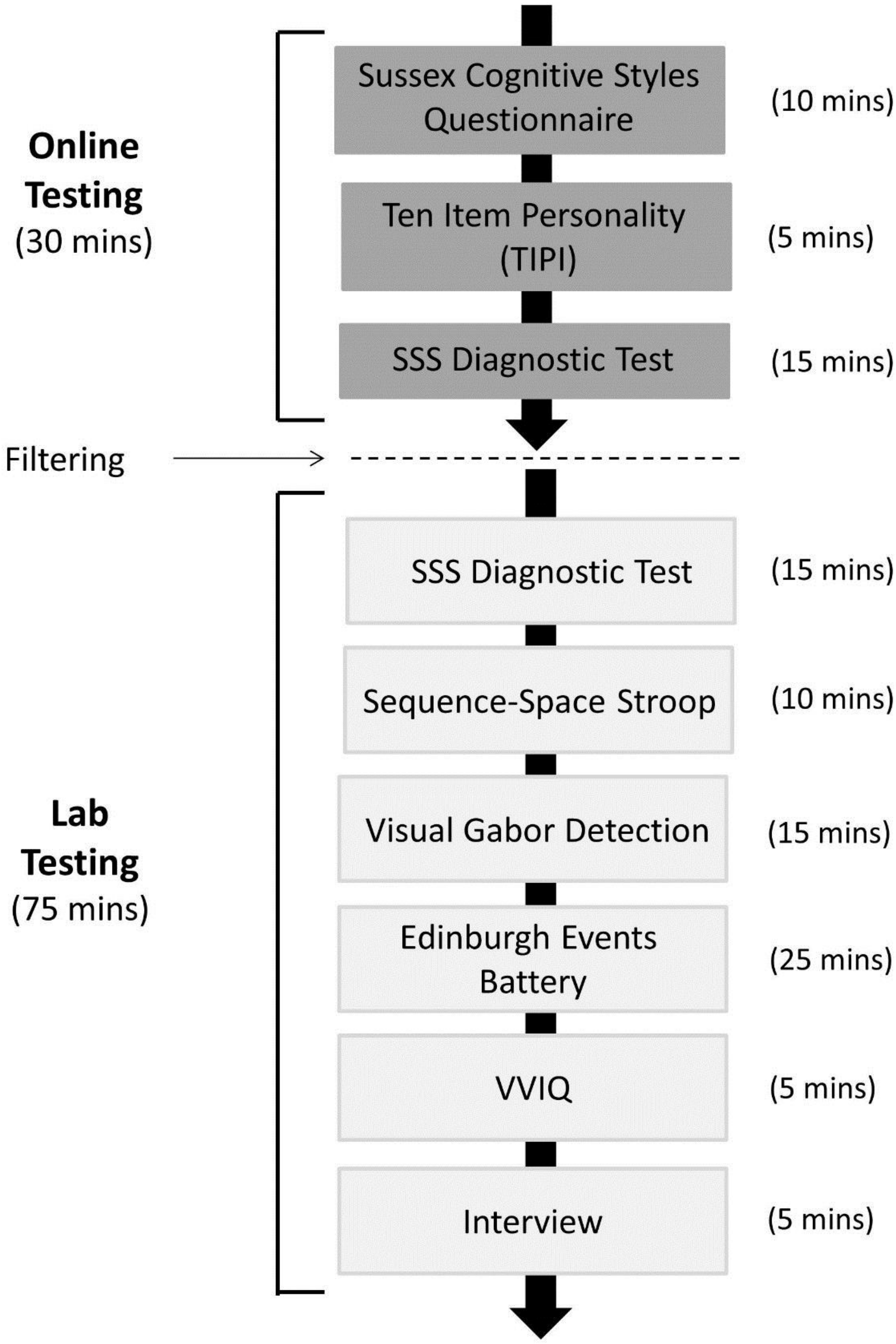


Figure 2: Timeline for Study 2.

<u>Session 1: Online Testing</u>

The online test was run via Qualtrics software (for the questionnaires) which then directed participants to Inquisit software for the spatial consistency test.

Sussex Cognitive Styles Questionnaire (Mealor et al., 2016). This is a 60 item questionnaire with each item answered on a 5 point Likert scale from ‘strongly disagree’ to ‘strongly agree’. The questions were presented in a random order but grouped into factors for analysis. The six factors with representative examples from each are: Imagery Ability (e.g. “My mental images are very vivid and photographic”), Technical/Spatial (e.g. “I can easily imagine and mentally rotate three-dimensional geometric figures”), Language and Word Forms (e.g. “When I hear a new word, I am curious to know how it is spelled”), Need for Organisation (e.g. “Order is important to me”), Global Bias (e.g. “I usually concentrate on the whole picture, rather than the small details”) and Systemising Tendency (e.g. “When I look at an animal, I like to know the precise species it belongs to”). ‘Systemising’ refers to a *motivation to understand* the rules and regularities that determine how a system functions, or the tendency to construct systems to understand the world whereas ‘Need for organisation’ relates to the *preference or need of order* in one’s environment or behaviour, for example for objects to be grouped according to common features, like shape or colour. The average response for each subscale was calculated (i.e. from 1 to 5), with reverse-coded questions being flipped as appropriate.

Ten Item Personality Inventory (TIPI) (Gosling, et al., 2003). This is a brief measure of the Big Five personality domains (two questions each for Openness to Experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism). Participants are presented with character traits (e.g. “Anxious, easily upset”) and asked to determine the extent to which it applies to themselves using a seven point Likert scale from ‘strongly disagree’ (1) to ‘strongly agree’ (7). The average response for each subscale was calculated (i.e. from 1 to 7), with reverse-coded questions being flipped as appropriate.

Spatial Consistency Test. This is identical to the procedure described in Study 1, except for one minor change. Specifically, each trial consisted of a central fixation cross (1 second) followed by a stimulus (e.g. “January”), which remained on the screen until a mouse click was made (the previous version presented the stimulus for 1 second).

Session 2: In-Person Testing

Spatial Consistency Test. This is similar to the procedure used in the first session with just a marginal change to the items and procedure. In our items, we continued to use months (N = 12), but replaced days with years (N = 7; "Year 1900", "Year 1918", "Year 1945", "Year 1980", "Year 2000", "Year 2010", "Year 2016") and took our number-stimuli from a wider numerical range (Experiment 1: 0-9; here 1, 10, 20, 30 (etc.) to 100). The main procedural difference was that the position of the stimulus was randomly jittered around the centre (between 45% and 55% of the screen width and height). This was to discourage participants from using this as an anchoring point and, to rely instead on their internal image. The test was added after testing the first batch of participants and, hence, has a smaller sample size (N=46).

Sequence-Space Stroop Test. All 64 participants from Session 2 (13 SSS, 51 Non-SSS) completed this task, and all SSS participants who took part in this task reported that they had a spatial arrangement for months. Data from three participants (1 SSS, 2 Non-SSS) were excluded due to poor task performance (see results). This was a novel test, albeit similar to others in the literature (e.g. Sagiv et al., 2006; Smilek et al., 2007). The test was run on Matlab, on a 39 x 29 cm CRT monitor (refresh rate of 85 Hz, colour depth of 24 Bit), at a viewing distance of 100 cm and in a dark room. The test was conducted using months as stimuli, and participants were initially asked to use a mouse click to locate months on the screen (as per the spatial consistency test). Each month was presented once only. These locations were then used as the congruent spatial coordinates in our Stroop task. Incongruent coordinates were created by rotating these coordinates 180 degrees through the centre of the screen. There were 12 month stimuli presented 5 times in each of its two possible locations and participants were required to make a speeded ‘button’ press (D or K) to indicate whether the month had been displayed on the left or right of the screen. In order to ensure that participants processed the meaning of the words and not just their location on the screen, there were 96 filler trials, made-up of 48 filler items consisting of nouns that resembled months orthographically (e.g. “Apron” / “April”). When non-month items appeared, participants had to withhold their normal response and press the space bar instead. Filler items were not analysed. Each trial began with a fixation dot (500 ms) followed by the stimulus (until the participant responded). All stimuli were presented in a white text, Calibri font size 25, on a mid-grey background. The task was preceded with a short practice block of 10 trials, where participants were presented with stimuli ‘Month’ or ‘Not month’ in order to familiarise themselves with the task.

Visual Gabor Detection. This task is identical to the one used by Ward et al. (2017) in which people with SSS had increased sensitivity at detecting low visibility stimuli that was particularly apparent for high spatial frequencies. The same monitor was used as in the previous Stroop task, during which time participants dark-adapted for the current task. Stimuli consisted of centrally presented vertical Gabor patches subtending a visual angle of 7.2°, with spatial frequency (SF) of 0.49 cycles/° (i.e., low spatial frequency, LSF) or 14 cycles/° (i.e., high spatial frequency, HSF). The Gabor patches were presented on a mid-grey monitor background (23 cd/m2), at 0.05 (low), 0.1 (medium), and 0.5 (high) contrast levels (Michelson contrast), considering the Gamma of the monitor. Participants were required to press the ‘space bar’ whenever they detected a stimulus on the screen. The task began with 8 practice trials, in which participants received feedback on correctness, followed by 5 experimental blocks containing no feedback. Each experimental block presented 5 high-contrast, 8 medium-contrast, and 10 low-contrast HSF stimuli, and 5 high, medium and low-contrast LSF stimuli. Four additional trials in which no stimulus was presented was also included, totalling 42 trials per block. Participants completed more HSF trials compared to LSF trials to ensure enough correct responses were made for reliable d-prime calculations, given the relative difficulty of these conditions. Each trial began with a central fixation cross, presented for a variable duration, randomly chosen from 500 ms to 1500 ms. The target stimulus was presented for 340 ms, overlaid on the fixation cross. After a response deadline of 1500 ms, the fixation cross disappeared and a 500 ms delay led to the beginning of the next trial.

The Edinburgh Events Battery. This is an updated version of the test reported in Simner et al. (2009) which investigates the ability to recall autobiographical events, and to place public events in time (i.e., state the year). The public events section was divided into two parts. The International World presented 60 international news events (e.g., World Trade Center attack) between 1950 and 2015 inclusive, with 10 events per decade. The second section was for Cultural Events specific to the UK and/or to English-language speakers and assessed memory of the years associated with films and songs respectively. The test of films included 30 movie titles for English-language films (mostly UK or US; e.g., The King's Speech) that won the Oscar for Best Picture between 1950 and 2015 (five films per decade). The popular music consisted of song names and artists (e.g., Bohemian Rhapsody; Queen) that were Number 1 Singles on Christmas day in the UK (five songs per decade between 1950-2015). Events were blocked according to type, and presented in random order within blocks. On each trial, the name of a single stimulus (event, film, song) appeared on screen above a response box. Participants typed the year they believed the event occurred, then hit 'Enter'. If no response was submitted after 6 seconds, the prompt "Quick!" appeared to warn participants that the end of the trial was approaching. After 10 seconds the trial timed out with a bell sound and the next stimulus was displayed on the screen. The first section of the public events battery (International World Events) is suitable for all of our international participants and the second section (UK/English Cultural Events) is suitable for our British subjects only. All subjects completed all sections but our analyses will reflect this nationality difference. The autobiographical section presented nine years within the life-time of each participant, equi-spaced between the year the participant was aged 5 and the current year minus 3. Years were presented in a random order for 90 seconds each, above 30 one-line empty text boxes. Participants were asked to type as many memories as possible about their life during the given year, using one response box per memory. At the end of the 90 seconds, the next year was displayed on screen and a bell informed participants that the next target year was displayed on-screen.

The final task of the events battery consisted of a typing speed control test (Warmington, Stothard, & Snowling, 2013), in which participants were required to type the sentence “Transportation is movement of people and goods from one location to another”, repeatedly for two minutes, from which the average words per minute were calculated. This was to allow us to ensure that differences in autobiographical-reporting had not been due to differences in typing speed. All tests in the events battery were run using WebExp2, a Java toolbox for web-based psychological experiments (Keller, Gunasekharan, Mayo, & Corley, 2009).

Vividness of Visual Imagery Questionnaire (VVIQ-2). The VVIQ-2 is a questionnaire involving 32 items (Marks, 1995), of which half are derived from the original VVIQ (Marks, 1973). Its aim is to assess the vividness of visual imagery. Each item is scored on a 5-point Likert scale, where 1 stands for no imagery present at all, and 5 - perfectly clear image, as vivid as normal vision (thus scores range from 32 to 160).

Interview. Participants were asked how they think about concepts such as time and number in their daily life, and how they went about choosing locations in the spatial consistency test. Follow-up questions asked about any mental images for numbers or time, their consistency over life, automaticity, location (mind’s eye, outside of body), and whether they remembering creating the spatial arrangement themselves. Each interview was audio recorded and then blind scored by different people to enable a calculation of inter-rater agreement. The scorers were the authors JW and EP, and an intern TC; both EP and TC have SSS themselves. The rating scale ranged from -5 to +5 in which +5 indicates certainty that the person has SSS, -5 indicates certainty that the person does not have SSS, and 0 indicates complete uncertainty as to their status.

Results

Spatial Consistency Tests

As in Study 1 participants were classed as having probable SSS if they had a consistency score <.203% obtained by non-trivial means (e.g. not clicking on the same location repeatedly), and a questionnaire score ≤19. Whereas in Study 1 we reassigned participants to the non-SSS group who obtained high consistency using trivial strategies by visual inspection, here we used an automated approach of checking that participants had clicked multiple parts of the screen. Namely we determined whether there was sufficient variability in the x-coordinates and/or y-coordinates (standard deviation>.075 for a normalised screen with width and height of 1 unit). The results are shown in Figure 3. The prevalence estimate from Study 2 was 12.8%. The prevalence in males was 10.8% (4/37) and in females it was 13.1% (26/198) giving a female: male ratio of 1.2:1 which was not statistically significant, $x^2(1) = 0.118$, $p = .731$.

	Consistency

		Pass	Fail
Qu'aire ≤ 19	Pass	12.8	10.1
	Fail	20.0	57.0

Figure 3: The estimated prevalence of SSS (grey), together with non-SSS groups, based on the University sample

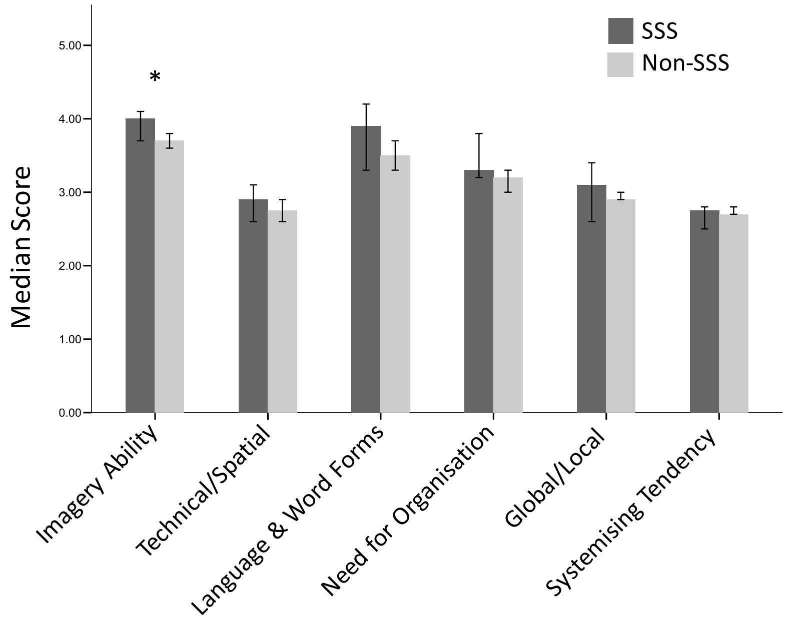
In all subsequent analyses, the three non-SSS participant groups are treated together as a single non-SSS category because treating them as three separate groups would greatly reduce our statistical power. However, a full breakdown is provided in the Supplementary Results for completeness.

The spatial consistency test performed in Session 2 was not an exact repeat of Session 1 (the stimuli for numbers were changed; days were replaced by years; and the stimulus position was jittered around the centre). . As such it does not present with a true measure of test-retest reliability. Nevertheless, the groupings are relatively stable even allowing for these differences. Considering all stimuli and the same stringent cut-off (consistency<.203, questionnaire≤19), 76% of participants retain the grouping they had at time 1. If one considers only months of the year (these being the only repeated stimuli) and a more liberal questionnaire cut-off (given that it asks about stimuli other than months) then the figure rises to 87% (consistency<.203, questionnaire≤21).

<u>Questionnaire Measures: SCSQ, TIPI, and VVIQ-2</u>

The results for the SCSQ and TIPI questionnaires are summarised in Figure 4, and are treated non-parametrically. Questionnaire data was missing from one participant (non-SSS). For the SCSQ, the only construct in which there was a significant group difference was Imagery Ability, *U = 1959.50, p = .006*, *r = .21, corrected for multiple tests*, with the SSS group (Mdn = 4.00) reporting significantly greater imagery ability than controls (Mdn = 3.70). Within the controls, those who passed the consistency test did not have higher imagery scores relative to those that failed, *U = 3480.50, p = .198, r = .09* suggesting that good performance on this objective measure is not simply due to high imagery. On the contrary, imagery was found to be higher in the controls who claim to have SSS on the questionnaire but fail the objective test *U = 852.50, p < .0005, r = .33*.

The Supplementary Results report the breakdown of scores amongst the Non-SSS group but it is to be noted here there imagery ability did not correlate with consistency score *rho*(232) = -.02, *p* = .821, or with the questionnaire score *rho*(232) = .12, *p* = .075, suggesting that it is unlikely that Imagery Ability is driving SSS status. For the TIPI, There were no significant differences between the two groups. As such, we feel confident that any large differences on the cognitive measures do not reflect group differences in personality.



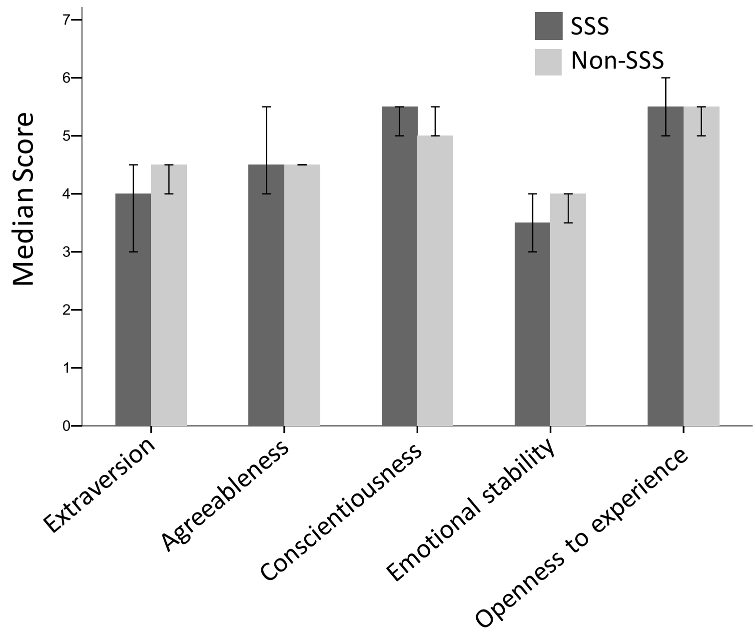


*Figure 4: Subscale median scores for the Sussex Cognitive Styles Questionnaire (top) and Ten Item Personality Inventory (bottom). Error bars show 95% CI, and * P<.05.*

<u>Sequence-Space Stroop Test</u>

All 64 subjects of Session 2 contributed to this test. Three participants (1 SSS, 2 Non-SSS) were excluded because of very high error rates indicating that they had not followed the instructions (accuracy was >2SD lower than the group mean). All SSS participants reported that they had spatial arrangements for months. We first analysed the percentage of correct responses using a 2 (SSS status) by 2 (congruency condition) mixed ANOVA. On average, both the SSS (M = 94.17, SD = 5.54) and the non-SSS group (M = 95.31, SD = 5.54) performed near-ceiling. There was no significant main effect of SSS status $F(1, 59) = 0.41$, $p = .526$, $\eta_p^2 = .01$, or Congruency $F(1, 59) = 2.51$, $p = .119$, $\eta_p^2 = .04$, but there was a significant interaction between the two $F(1, 59) = 4.85$, $p = .032$, $\eta_p^2 = .04$. Descriptive statistics suggest that in the SSS group, the difference between accuracy in the congruent (M = 96.67, SD = 5.37) and incongruent (SD = 91.67, SD = 9.37) conditions was larger than that in the non-SSS group (congruent M = 94.90, SD = 7.25; incongruent M = 95.71, SD = 6.12). However, a breakdown of this interaction, examining the effect of congruency as a function of group did not reach significance, SSS: $t(11) = 1.56$, $p = .146$, *Cohen's d* = .61; non-SSS: $t(48) =-0.77$, $p = .443$, *Cohen's d* = .12.

Only correct trials were included in the RT analysis. To obtain a measure of congruency effect magnitude, the difference in RT for correct trials between the congruent and incongruent condition was divided by the sum of RTs. Here, more positive values indicate that the congruent condition was responded to faster than the incongruent condition. In other words, the more positive the value, the more interference there was resulting from the presentation of a month in an incongruent location to that originally placed by the participant. The congruency effect was larger in the SSS (M = .05, SD = .08) compared to the Non-SSS group (M <.01, SD = .06), $t(59) = 2.27$, $p = .027$, *Cohen's d* = .66.

In conclusion, this test provides support for one of the key defining features of sequence-space synaesthesia, namely automaticity. Participants with synaesthesia show interference from their spatial form even when it is irrelevant to the task. The mechanism behind automatic effects in synaesthesia remains debated: for instance, it has been suggested that it may just reflect over-learning (Price & Mattingley, 2013). However, for present purposes, we merely wish to note that those who we diagnose with SSS behave differently on this measure, as noted before (e.g. Smilek et al., 2007).

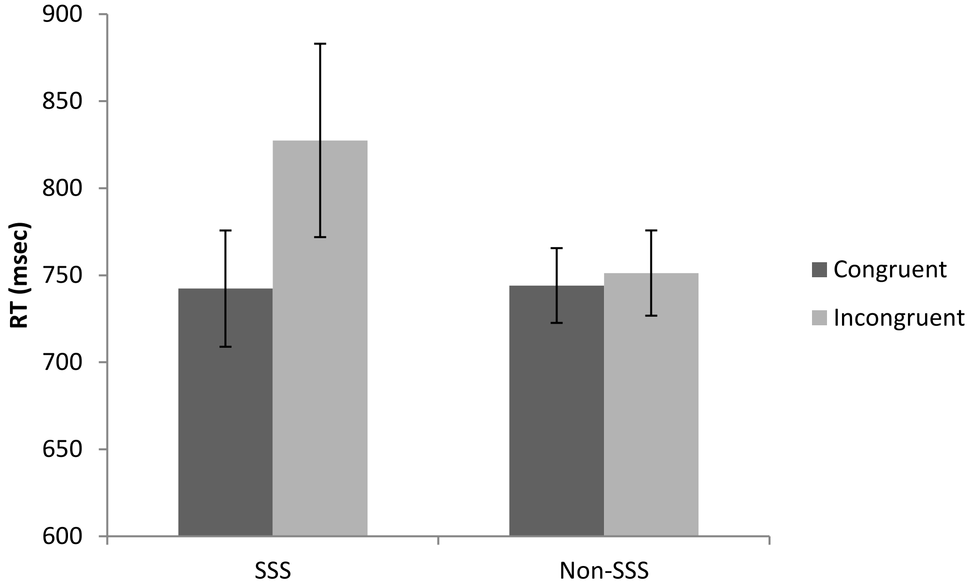


Figure 5: Response times to correctly identify the left/right spatial location of months on the screen as a function of whether they are congruent or incongruent with their own initial location selections. Error bars show +/- 1 SEM.

<u>Visual Gabor Detection</u>

Participants were required to make a button press when a visual stimulus was present and withhold a response when it was absent. Four participants failed to detect any high spatial frequency targets whatsoever and so were excluded from the analysis (1 SSS, 3 Non-SSS), and two (1 SSS 1 Non-SSS) failed to complete the task due to technical errors. For the remaining participants, d-prime was calculated from the hits (visual stimuli reported as 'seen') and false alarms (null trials reported as 'seen'). The results are summarised in Figure 6 (and Supplementary Material).

The results were analysed as a 2x2x3 ANOVA contrasting group, spatial frequency (high/low), and contrast (high/medium/low). There were significant main effects of both spatial frequency (LSF being easier) $F(1,56) = 127.72$, $p < .001$, $\eta_p^2 = .70$, and contrast (high contrast being easier), $F(2,112) = 137.48$, $p<.001$, $\eta_p^2 = .71$, as well as an interaction between the two (contrast exerting a bigger influence on the harder HSF stimuli) $F(2,112) = 63.86$, $p<.001$, $\eta_p^2 = .53$. There was also a significant main effect of group (SSS perform better) $F(1,56) = 14.74$, $p<.001$, $\eta_p^2 = .21$, and a significant three way interaction between group, spatial frequency and contrast ($F(2,112) = 3.83$ $p = .025$. $\eta_p^2 = .07$). This is due to some stimuli being easier to detect by the SSS group than others, although we are cautious about interpreting the interaction given that performance was close to ceiling on low spatial frequency stimuli. No other interaction approached significance (all p's > .10).

The large effect sizes observed are likely to constitute genuine differences in visual ability given that: detecting low visibility stimuli is unlikely to be improved by being more motivated; having SSS can't be used strategically to assist performance on this task; the SSS group had been recruited in the same way as the controls and had not been singled out as 'special' during either recruitment or testing.

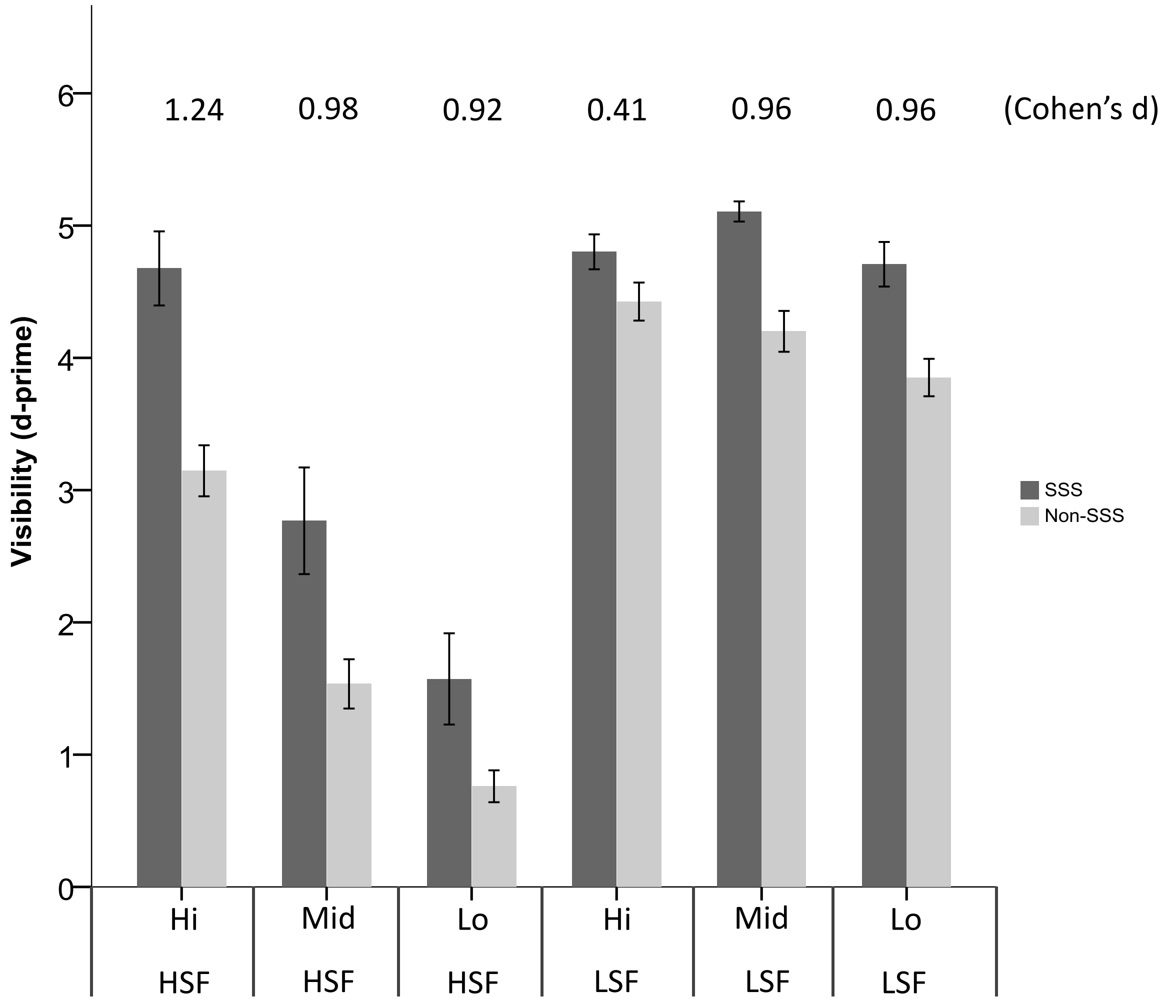


Figure 6: The ability to detect low visibility stimuli (d-prime) depending on spatial frequency (HSF=high, LSF=low) and contrast (hi[gh], mid[dle], lo[w]). Error bars show +/- 1 SEM.

<u>The Edinburgh Events Battery</u>

Public Events. Our dependent measures were accuracy and reaction time. As in the original study, our accuracy measure for each subject was an error-score representing the mean distance in years between the correct event date and the participant's response. Also as in the original study we began by removing all responses that fell outside the range of dates specified in the task instructions (1950-2015) and cleaned the data by removing outliers >3SD from the mean distance for each individual. The battery had previously been validated on an older sample (Simner et al., 2009) and our younger sample noted that many of the earlier events were largely unfamiliar (and were generally dated inaccurately). Hence, we focussed on events from within their own lifetime (the years 1996 onwards were used). For the international section of the test (International News Events) we included all 63 international participants. For the UK-relevant section (UK/English Culture: film and songs) we include all UK subjects (N = 10 SSS group; N = 23 non-controls). Given the overlap in subject-groups we corrected our p-values for multiple comparisons.

The SSS-group were significantly more accurate than controls in dating International News events. An independent samples t-test showed that synaesthetes' responses (M = 5.37, SD = 2.45) were on average significantly closer to the correct event date, compared to those of non-synaesthetes (M = 8.22, SD = 4.60), $t(41.02) = -3.06$, $p = .012$, *Cohen's d* = .77, *two-tailed, corrected for unequal variance; corrected for multiple comparisons across the three event-types*. We also determined that synaesthetes were not better simply by being slower. There was no significant difference between in response times between synaesthetes (M = 5455, SD = 1142) and non-synaesthetes (M = 5455, SD = 768), $t(61) = -0.22$, $p = .828$ (two-tailed, uncorrected). These data are presented in Figure 7.

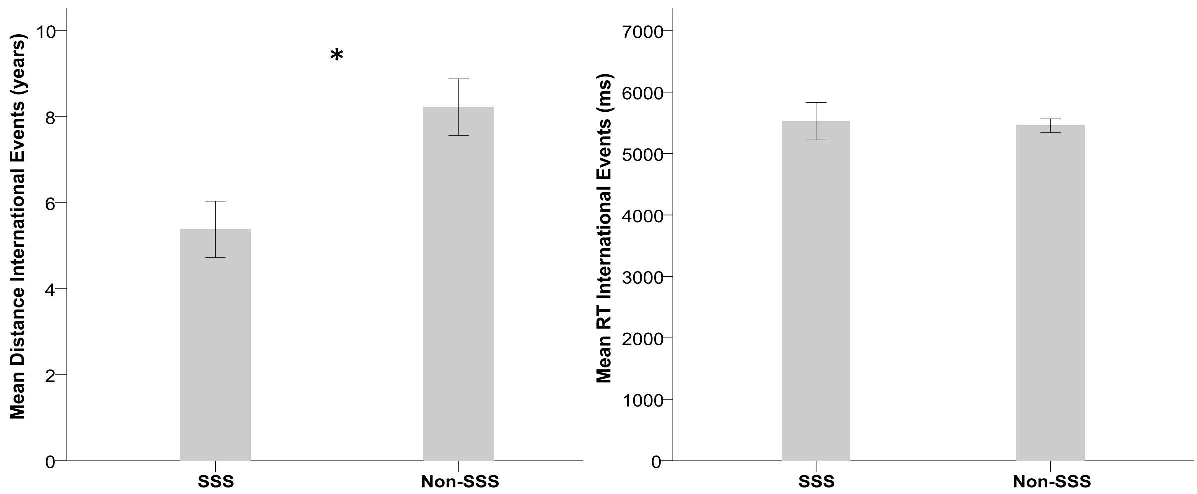


*Figure 7: Left panel shows mean distance (in years) between the correct year of international events and participant responses as a function of synaesthete status (N = 63 international participants). Note that smaller bars mean greater accuracy; error bars represent -/+ 1 SEM. * p = .011. Right panel shows mean reaction times.*

The smaller, British-only SSS-group were not significantly more accurate than controls in dating UK/English-language Cultural Events (films; songs). A 2x2 mixed ANOVA crossing Event type (Films, Songs) and Synaesthete status (SSS, non-SSS) showed no main effects of event type, $F(1,33) = 0.79$, p = .380, $\eta_p^2 = .02$), or group, $F(1,33) = .01$, $p = .932$, $\eta_p^2 < .001$) and there was no significant interaction, $F(1,33) = .67$, p = .672, $\eta_p^2 = .02$. We also ran a 2 x 2 ANOVA, this time on reaction times. Although there was a significant effect of event type $F(1,33) = 16.68$, p < .0005, $\eta_p^2 = .34$), with reaction times for films (M = 4292, SD = 1090) faster than for songs (M = 4873 , SD =1116), there was no effect of group ($F(1, 33)$ 2.59) = .071, $p = .792$, $\eta_p^2 = .002$) and no interaction ($F(1, 33) = 2.59$, $p = .117$, $\eta_p^2 = .07$).

Autobiographical Events. For the autobiographical events, the number of memories recalled did not meet assumption of normality for either group so we used non-parametric tests. There was no significant difference in the number of memories between the SSS group (Mdn = 48.00) and the non-SSS group (Mdn = 45.00) U = 285.50, $p = .342$. There was also no difference in the detail of the memories, considering the average number of words given per event (SSS M = 4.52, SD = 1.54; non-SSS M = 5.10, SD = 1.40), $t(61) = 1.36$, $p = .180$, Cohen's d = 0.39, 95%CI [-0.27, 1.45]. On the control typing task, there was no significant difference between groups in typing speed, although there was a trend for the SSS group to type more words per minute (SSS M = 54.6, SD = 11.36; Non-SSS M = 47.10, SD = 14.23), $t(61) = 1.82$, $p = .073$; *Cohen's d* = 0.58, 95%CI [-15.75, 0.73].

Interview Classifications

One participant did not participate in the interview (i.e. N = 63). Three scorers, all blind to interviewee status, each gave a score between -5 (certain interviewee does not have SSS) and +5 (certain interviewee does have SSS) based on audio recordings of the interview. Inter-rater score correlations were very high, indicating agreement across raters regarding participants' synaesthete status (r = 0.94, 0.97 and 0.77, respectively). The scores were averaged across the raters to provide a single interview score. On average, the SSS group (M = 2.54, SD = 3.80) scored significantly higher in the interview, than the Non-SSS group (M = -0.33, SD = 3.81), $t(21.06)=2.51$, $p = .020$, corrected for unequal variance; Cohen's d = .75, 95%CI [0.50, 5.24].

Eleven of the 13 participants in the SSS group had been given a positive score by the raters, but two had been rated as non-SSS. We looked at the profile of the two discrepant cases more carefully on the tasks shown to be most sensitive to SSS (i.e. Stroop and visual perception). One participant had only weak evidence of a congruency effect on the Spatial Stroop (696 v. 725 msec for congruent and incongruent trials) and was the only SSS participant who had to be excluded from the visual Gabor detection task. The other participant had performed at the high level of other SSS participants on the visual Gabor detection task (d-prime scores for HSF of 4.3, 4.5 and 2.9 for high, medium and low contrast) but was the only SSS participant to be excluded from the Stroop test. Thus, we can't make firm diagnostic conclusions about these two individuals perhaps because the quality of their data (in terms of compliance with instructions) was patchy overall.

Conversely, the raters expressed confidence in N = 15 Non-SSS participants having SSS (mean score > 3). Six of these 15 individuals belonged to the group that had previously self-declared SSS on the questionnaire but failed the consistency test, 5 were in the group that passed the consistency test but didn't self-declare and the remaining 4 failed both tests. The performance of this N = 15 sub-group on the visual detection and Stroop tasks was unremarkable; their average d-prime for the high, medium and low contrast within the HSF condition was 3.26, 1.91 and 0.93, respectively; and the difference between the RTs of the congruent and incongruent Stroop conditions for this group was not significant (congruent 715 [SD=200], incongruent = 755 [SD=257]; paired t(12)=1.286, p=.223, Cohen's d=.343).

Summary

In summary, although we cannot be certain that our diagnostic test of SSS successfully categorises all participants it fairs better than classification based on interviews (including by an expert on synaesthesia) in terms of predictive validity on other tasks. Table 2 summarises the relationship between the various key tasks and measures. Presence/absence of SSS, as determined by our new measure, was the only significant correlate of performance. Alternative measures of SSS (e.g. interview scores) and potentially confounding variables (e.g. imagery ability) did not yield any significant associations.

Table 2. Pearson's correlations between presence of SSS (coded as 0 or 1), interview score (-5 to 5), visual detection (mean d-prime for the 6 conditions), spatial Stroop interference (incongruent – congruent RTs divided by summed RTs), and International events (where a lower score implies a closer date estimate).

	Interview score	Imagery Ability	Visual detection	Stroop interference	International events (Mean distance)
SSS v. Non_SSS	**.284 (.024) (N=63)**	**.319 (.011) (N=234)**	**.457 (<.001) (N=58)**	**.283 (.027) (N=61)**	**-.273 (.031) (N=63)**
Interview score		.166 (.198) (N=62)	.197 (.142) (N=57)	.242 (.063) (N=60)	-.205 (.119) (N=59)
Imagery ability			.110 (.417) (N=57)	.184 (.160) (N=60)	-.140 (.277) N=62)

Visual detection				.094 (.495) (N = 55)	-.165 (.233) (N=54)
Stroop interference					.165 (.219) (N = 57)

General Discussion

The aim of this study was to extend a recently devised diagnostic test for sequence-space synaesthesia to naïve samples who had not come to our attention by self-referring as synaesthetes. Our aim was to estimate the prevalence and the cognitive abilities linked to this phenomenon. In order to do this we devised a diagnostic test for SSS based on a consistency test from Rothen et al. (2016) and including a new questionnaire measure. To be classed as having SSS, participants had to meet both an objective criterion on the spatial consistency test, and also had to report the subjective presence of SSS in daily life using our questionnaire. We first validated our test against a sample of people likely to possess SSS (recruited from the synaesthesia community) and we then estimated the prevalence of SSS in a naïve sample to be 8.1% (Study 1) and 12.8% (Study 2). In Study 2 we invited participants back for more detailed cognitive testing and showed that those with SSS reported higher imagery ability (on SCSQ, Mealor et al., 2016), showed stronger interference effects on spatial Stroop test, had significantly enhanced abilities on a test of visual perception (detecting low visibility stimuli), and had significantly better memory for dating the year of International news events. We did not find better memory for recalling autobiographical events nor when dating UK-oriented cultural events -- although this latter was in a far smaller sample of UK-only subjects. We discuss these findings and their wider implications below.

Our two studies yielded somewhat different estimates of prevalence, although both figures are consistent with previous estimates based on more informal methods. We are inclined to give more weight to the 8.1% estimate because it was derived from a larger sample that is likely to demographically representative. It is conceivable that people with unusual experiences (such as SSS) gravitate more towards psychology as a discipline which may inflate the estimate in Study 2 (previous studies have shown this is true for the prevalence of grapheme-colour synaesthesia amongst arts students; Rothen & Meier, 2010). As with other types of synaesthesia (Simner & Carmichael, 2015) we did not find a significant difference in prevalence across sexes. It is also to be noted that our prevalence estimate is for multiple kinds of sequence-space synaesthesia, which we think is typical of this phenomenon, but our test may not adequately capture those with single spatial forms (i.e. it is a lower estimate of prevalence).

To some extent it is not possible to know how good this particular diagnostic test is because there is no independent yardstick for determining who does and does not have SSS. However, its validity comes from examining converging evidence on independent tests. There was significant agreement between the diagnostic test and blind scoring of interviews with participants (done by a synaesthesia researcher [JW], and two students with SSS themselves). However, the agreement was not perfect: the interview ratings tended to give higher estimates of the number of people with SSS. The new diagnostic test is to be favoured over the interview ratings for two reasons. Firstly, it can be more easily reproduced across labs and, secondly, because it more closely tracked performance on other independent measures.

The study also sheds light on the nature of sequence-space synaesthesia itself. SSS appears to be related to increased imagery ability as first conjectured by Galton (1880a, 1880c). This was found on the Imagery Ability subscale of the SCSQ (Mealor et al., 2016), which is closely related to the 'object imagery' subscale of the OSIQ (Blazhenkova & Kozhevnikov, 2009). We didn't find a difference in the 'Technical / Spatial' factor which, amongst other things, relates to the manipulation of mental images and this fits with previous literature (e.g., Rizza & Price, 2012). We found a non-significant trend for VVIQ-2 which, compared to later measures, focusses only on vividness and not additionally on habitual use of imagery. Although SSS is related to mental imagery, it cannot be reduced to merely being good at mental imagery. Mental imagery scores vary greatly amongst the Non-SSS group. Those within the Non-SSS group who passed the consistency test (i.e. because they generated a plausible and consistent spatial form on demand) were not the ones with the highest imagery, as we initially predicted. If anything, it is those who <u>fail</u> the consistency test but claim to have SSS (on the questionnaire) who have more vivid mental imagery (on VVIQ and SCSQ). These individuals also tend to be classified, at interview, as candidates for having SSS perhaps because of their proneness to use visuo-spatial thinking.

High mental imagery may be necessary for the development of SSS, as suggested by others (e.g. Price & Pearson, 2013), although it is unlikely to be sufficient. Our research also highlights a wider profile of differences. In the studies reported here the only large effect sizes (Cohen's $d>.80$) that we found were on the test of visual perception involving detection of Gabor gratings varying in their visibility (due to contrast and spatial frequency manipulations). This replicates the findings of Ward et al. (2017) on the same visual test (previously using a self-referred SSS group). Strictly speaking, this test does not show that SSS is a visual phenomenon. What it does suggest is that some aspects of the visual system in people with SSS are abnormally sensitive. Spatial frequency tuning is only present in cortical neurons (area V1 and above), and the processing of high spatial frequencies (for processing fine detail) is part of the parvocellular system which is also colour-selective (Maunsell, 1987). We do not know whether the presence of SSS causes these changes in the visual system, or whether changes in the visual system are a precursor to developing SSS. The time frame in which spatial vision matures, up to 7 years (Ellemberg, Lewis, Liu, & Maurer, 1999), coincides with the period in which children learn these sequences. We also note that people with synaesthesia, including SSS, report differences in subjective sensory sensitivity (e.g. aversion to certain lights, sounds) (Ward, Brown, Sherwood, & Simner, in press).

Some have suggested that other kinds of synaesthesia reflect memory of stimuli in their environment, for example exposure to coloured letters for grapheme-colour synaesthetes (Witthoft, Winawer, & Eagleman, 2015; Yon & Press, 2014). Galton's claim was that spatial forms were constructed by the child from the spoken word sequences and subsequently remembered or forgotten, rather than learned from the environment. Although we cannot discount a role of either of these learning and memory processes, the findings from tests of visual perception and mental imagery make it unlikely that SSS is <u>just</u> a memory phenomenon. But still, memory was superior in our study too, at least in some tasks. Although not an episodic memory test (which has often been used in synaesthesia, Rothen, Meier, & Ward, 2012) our task dating International News Events showed a significant advantage for the SSS group (replicating Simner et al., 2009). There was no effect when dating UK-oriented Cultural events, although there were notably fewer subjects in this test given the requirement for UK-subjects only. There was also no superiority recalling autobiographical events from a given year, contrary to effects found previously (Simner et al., 2009). This benefit may directly tap spatial forms for years in particular (which were the group selected by Simner et al.) or the autobiographical advantage may be more apparent in older synaesthetes (the mean age in Simner et al. was 36 years).

In summary, we extend and validate a new diagnostic test of sequence-space synaesthesia. We provide an estimate of its prevalence in two samples, and we show that it is linked to individual differences in cognitive ability. Our results also shed new light on the nature of this interesting and important phenomenon.

<u>Acknowledgements</u>

The research leading to these results received funding for author JS from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. [617678].

<u>References</u>

Banissy, M. J., Holle, H., Cassell, J., Annett, L., Tsakanikos, E., Walsh, V., … Ward, J. (2013). Personality traits in people with synaesthesia: Do synaesthetes have an atypical personality profile? Personality and Individual Differences, 54(7), 828–831. http://doi.org/10.1016/j.paid.2012.12.018

Blazhenkova, O., & Kozhevnikov, M. (2009). The New Object-Spatial-Verbal Cognitive Style Model: Theory and Measurement. *Applied Cognitive Psychology, 23*(5), 638-663. doi: 10.1002/acp.1473

Brang, D., Miller, L.E., McQuire, M., Ramachandran, V. S., & Coulson, S. (2010). Enhanced mental rotation ability in time-space synesthesia. *Cognitive Processes, 14*(4), 429-34. doi: 10.1007/s10339-013-0561-5

Brang, D., Teuscher, U., Ramachandran, V. S., & Coulson, S. (2010). Temporal sequences, synesthetic mappings, and cultural biases: The geography of time. *Consciousness and Cognition, 19*(1), 311-320. doi: 10.1016/j.concog.2010.01.003

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science, 6*(1), 3-5. doi: 10.1177/1745691610393980

Burbridge, D. (1994). Galton's 100: An exploration of Francis Galton's imagery studies. *British Journal for the History of Science, 27*, 443-463.

Calkins, M. W. (1895). Synaesthesia. *American Journal of Psychology, 7*, 90-107.

Chun, C. A., & Hupé, J. M. (2016). Are synesthetes exceptional beyond their synesthetic associations? A systematic comparison of creativity, personality, cognition, and mental imagery in synesthetes and controls. British Journal of Psychology, 107(3), 397–418. http://doi.org/10.1111/bjop.12146

Eagleman, D. M. (2009). The objectification of overlearned sequences: A new view of spatial sequence synesthesia. *Cortex, 45*(10), 1266-1277. doi: 10.1016/j.cortex.2009.06.012

Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D., & Sarma, A. K. (2007). A standardized test battery for the study of synesthesia. *Journal of neuroscience methods, 159*, 139-145.

Ellemberg, D., Lewis, T. L., Liu, C. H., & Maurer, D. (1999). Development of spatial and temporal vision during childhood. *Vision Research, 39*(14), 2325-2333. doi: 10.1016/s0042-6989(98)00280-6

Flournoy, T. (1893). *Des phemomenes de synopsie*. Paris (maybe Geneva???): Alcan.

Galton, F. (1880a). Statistics on mental imagery. *Mind, 5*, 301-318.

Galton, F. (1880b). Visualised numerals. *Journal of the Anthropological Institute, 10*, 85-102.

Galton, F. (1880c). Visualised numerals. *Nature, 21*, 494-495.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A Very Brief Measure of the Big Five Personality Domains. *Journal of Research in Personality, 37*, 504-528.

Havlik, A. M., Carmichael, D. A., & Simner, J. (2015). Do sequence-space synaesthetes have better spatial imagery skills? Yes, but there are individual differences. *Cognitive processing, 16*(3), 245-253. doi: 10.1007/s10339-015-0657-1

Hubbard, E. M., Ranzini, M., Piazza, M., & Dehaene, S. (2009). What information is critical to elicit interference in number-form synaesthesia? *Cortex, 45*(10), 1200-1216. doi: 10.1016/j.cortex.2009.06.011

Jewanski, J., Day, S. A., Simner, J., & Ward, J. (2011). The Development of a Scientific Understanding of Synesthesia from Early Case Studies (1849–1873). *Journal of the History of Neurosciences, 20*, 284-305.

Keller, F., Gunasekharan, S., Mayo, N., & Corley, M. (2009). Timing accuracy of Web experiments: a case study using the WebExp software package. *Behavioural Research Methods, 41*, 1-12.

Maunsell, J. H. R. (1987). Physiological evidence for two visual subsystems. In L. M. Vaina (Ed.), *Matters of Intelligence*. Dordrecht: Reidel.

Mealor, A. D., Simner, J., Rothen, N., Carmichael, D. A., & Ward, J. (2016). Different Dimensions of Cognitive Style in Typical and Atypical Cognition: New Evidence and a New Measurement Tool. *PLoS One, 11*(5), e0155483.

Patrick, G. T. W. (1893). Number-forms. *Popular Science Monthly, February*, 506.

Phillips, D. E. (1896-97). Genesis of number-forms. *American Journal of Psychology*, 506-527.

Price, M. C. (2009). Spatial forms and mental imagery. *Cortex, 45*(10), 1229-1245. doi: 10.1016/j.cortex.2009.06.013

Price, M. C., & Mattingley, J. B. (2013). Automaticity in sequence-space synaesthesia: A critical appraisal of the evidence. *Cortex, 49(5)*, 1165-1186. doi: 10.1016/j.cortex.2012.10.013

Price, M. C., & Pearson, D. G. (2013). Toward a visuospatial developmental account of sequence-space synesthesia. *Front Hum Neurosci, 7*

Ramachandran, V. S., & Hubbard, E. M. (2003). The phenomenology of synaesthesia. *Journal of Consciousness Studies, 10*, 49-57.

Rizza, A. & Price, M.C. (2012). Do sequence-space synaesthetes have better spatial imagery skills? Maybe not. *Cognitive Processes*, 13 Suppl 1, S299-303. doi: 10.1007/s10339-012-0459-7.

Rothen, N., Jünemann, K., Mealor, A. D., Burckhardt, V., & Ward, J. (2016). The sensitivity and specificity of a diagnostic test of sequence-space synaesthesia. *Behavior Research Methods, 48*(4), 1476-1481.

Rothen, N., & Meier, B. (2010). Higher prevalence of synaesthesia in art students. *Perception, 39*(5), 718-720. doi: 10.1068/p6680

Rothen, N., Meier, B., & Ward, J. (2012). Enhanced memory: Insights from Synaesthesia. *Neuroscience and Biobehavioral Review, 36*(8), 1952-1963.

Rothen, N., Seth, A. K., Witzel, C., & Ward, J. (2013). Diagnosing synaesthesia with online colour pickers: maximising sensitivity and specificity. *Journal of neuroscience methods, 215*(1), 156-160. doi: 10.1016/j.jneumeth.2013.02.009

Rouw, R., & Scholte, H. S. (2016). Personality and cognitive profiles of a general synesthetic trait. Neuropsychologia, 88, 35–48. http://doi.org/10.1016/j.neuropsychologia.2016.01.006

Sagiv, N., Simner, J., Collins, J., Butterworth, B., & Ward, J. (2006). What is the relationship between synaesthesia and visuo-spatial number forms? *Cognition, 101*, 114-128.

Seron, X., Pesenti, M., Noel, M.-P., Deloche, G., & Cornet, J. A. (1992). Images of numbers, or "when 98 is upper left and 6 sky blue". *Cognition, 44*, 159-196.

Shriki, O., Sadeh, Y., & Ward, J. (2016). The emergence of synaesthesia in a neural network model via changes in perceptual sensitivity and plasticity. *PLoS Computational Biology, 12*(7), e1004959.

Simner, J., & Carmichael, D. A. (2015). Is synaesthesia a dominantly female trait? *Cognitive Neuroscience, 6*(2-3), 68-76. doi: 10.1080/17588928.2015.1019441

Simner, J., Mayo, N., & Spiller, M.-J. (2009). A foundation for savantism? Visuo-spatial synaesthetes present with cognitive benefits. *Cortex, 45*(10), 1246-1260. doi: 10.1016/j.cortex.2009.07.007

Smilek, D., Callejas, A., Merikle, P., & Dixon, M. (2007). Ovals of time: Space-time synesthesia. *Consciousness and Cognition*. *16*(2), 507-519.

Spiller, M. J., Jonas, C. N., Simner, J., & Jansari, A. (2015). Beyond visual imagery: How modality-specific is enhanced mental imagery in synesthesia? *Consciousness and Cognition, 31*, 73-85. doi: 10.1016/j.concog.2014.10.010

Ward, J., Brown, P., Sherwood, J., & Simner, J. (in press). An Autistic-Like Profile of Attention and Perception in Synaesthesia. Cortex

Ward, J., Rothen, N., Chang, A., & Kanai, R. (2017). The structure of inter-individual differences in visual ability: Evidence from the general population and synaesthesia. *Vision Research*, 141, 293-302.

Warmington, M., Stothard, S. E., & Snowling, M. J. (2013). Assessing dyslexia in higher education: The York adult assessment battery-revised. *Journal of Research in Special Educational Needs, 13*, 48-56.

Witthoft, N., Winawer, J., & Eagleman, D. M. (2015). Prevalence of Learned Grapheme-Color Pairings in a Large Online Sample of Synesthetes. *PLoS One, 10*(3). doi: 10.1371/journal.pone.0118996

Yon, D., & Press, C. (2014). Back to the future: synaesthesia could be due to associative learning. *Frontiers in psychology, 5*. doi: 10.3389/fpsyg.2014.00702

Appendix

Note the questions were arranged on 5 pages: page 1 contained Q1 and Q2; page 2 contained the example spatial forms and Q3; Page 3 contained Q4-Q6; Page 4 contained Q7-Q10; and Page 5 contained Q10-15.

1. What is your age?

2. Please select your gender (1= female, 2= male, 3=undisclosed)

3. Some people routinely think about sequences as arranged in a particular spatial configuration, do you think this might apply to you? (1=strong agree, 5= strongly disagree)

4. Which of the following do you routinely think about in this way? (Numbers, days, months, years, letters of the alphabet, temperature, height, weight)

5. Where do you tend to routinely experience these sequences? (1= in the space outside my body; 2= on an imagined space that has no real location; 3= inside my body; 4= this doesn't apply to me!)

6. What kind of characteristics do these spatial sequences always tend to take? (Colors; shading; 2D; 3D; perspective; like blocks or tiles; a certain font)

7. Before doing this experiment, I always thought about NUMBERS as existing in a particular spatial sequence (1= strongly agree; 5= strongly disagree)

8. Before doing this experiment, I always thought about DAYS OF THE WEEK as existing in a particular spatial sequence (1= strongly agree; 5= strongly disagree)

9. Before doing this experiment, I always thought about MONTHS OF THE YEAR as existing in a particular spatial sequence (1= strongly agree; 5= strongly disagree)

10. I use this way of thinking about spatial sequences in my everyday life (1= strongly agree; 2= strongly disagree)

11. When doing the experiment, I didn't have any strong intuition as to where to put the NUMBERS (1= strongly agree; 5= strongly disagree) [Reverse coded]

12. When doing the experiment, I didn't have any strong intuition as to where to put the DAYS OF THE WEEK (1= strongly agree; 5= strongly disagree) [Reverse coded]

13. When doing the experiment, I didn't have any strong intuition as to where to put the MONTHS OF THE YEAR (1= strongly agree; 5= strongly disagree) [Reverse coded]

14. This experiment didn't really make much sense to me (1= strongly agree, 5= strongly disagree) [Reverse coded]

15. Feel free to enter any comments here. E.g. what strategy did you use? Do you want to clarify any of the above answers?